DOCUMENT RESUME

ED 160 660 95 TM 007 963

AUTHOR Cooley, William W.
TITLE Explanatory Observational Studies.
PUB DATE 28 Mar 78
NOTE 23p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 27-31, 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Data Analysis; Educational Research; *Hypothesis Testing; Models; *Observation; Program Effectiveness; *Research Design; *Research Needs; Sampling; State of the Art Reviews; Statistical Analysis

ABSTRACT

The state of the art in observational studies is discussed, with emphasis on explanatory research--studies of program effectiveness in which an attempt is made to explain the reasons for the results as well as describing the results. In addition to the collection of data and presentation of statistical facts, the explanatory study involves the specification of causal hypotheses, estimation of causal models, and testing of the validity of these models and hypotheses. Quantitative observational studies, when conducted according to the author's recommendations, can yield consistent, convincing, and useful explanations of educational processes. Three critical features of such studies are specified. (1) The definition of the sample and the populations studied affect the generalizability of the observed relationships. (2) Specification errors may result when the theoretical model describing the hypothesized causal structure of the variables under consideration is invalid; such problems result from subject matter rather than statistical errors. (3) The statistical procedures should reflect the multidimensional domains of educational treatments. Commonality analysis and a computer program, LISREL, are recommended as useful ways to identify the roles and reliabilities of the variables of interest. (Author/GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

William W. Cooley

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

ED160660

TM007 963

EXPLANATORY OBSERVATIONAL STUDIES

William W. Cooley

Learning Research and Development Center

and

School of Education

University of Pittsburgh

One does not normally dedicate a paper, but I would like to dedicate this one to Paul Lazarsfeld. The Learning Research and Development Center benefited greatly from his presence during the final years of his very distinguished career.

This paper was the basis for an American Educational Research Association Invited "State of the Art" Address, Toronto, Canada, March 28, 1978. Previous drafts benefited greatly from discussion with Paul Lohnes and Gaea Leinhardt. I am also indebted to James Algina, Lloyd Bond, and David Wiley for "forcing" me to examine the potential of structural equations. Additional helpful suggestions have been received from Bill Bickel, Betty Boston, Tony Eichelberger, Connie Faddis, Evelyn Fisher, Bill Gephart, Bob Glaser, Leo Klopfer, Maury Lindvall, Sherryl May, Charles Perfetti, Lee Poynor, Jiri Nehnevajsa, Charlie Schulberg, and Jerry Waller. Or, as Lennon and McCartney (1967) once put it, I had "a little help from my friends." Of course none of them can be held responsible for the weaknesses still present in this version.

The research reported herein was supported by the Learning Research and Development Center, supported in part as a research and development center by funds from the National Institute of Education (NIE), United States Department of Health, Education, and Welfare. The opinions expressed do not necessarily reflect the position or policy of NIE, and no official endorsement should be inferred.

EXPLANATORY OBSERVATIONAL STUDIES

William W. Cooley

The task assigned to me for today was to describe the state of the art in the design and analysis of research studies involving relationships among variables. That seemed impossible to do in twenty-five minutes, so a necessary first step was to decide on a particular focus.

The focus of these remarks is on what the statistical literature refers to as observational studies. More specifically, studies that inquire into the learning and development of human beings in the natural environments in which these processes occur. These studies are multivariate and longitudinal, recording the variances and covariances of events and individual differences as they occur and unfold. I will particularly focus on observational studies that are intended to be explanatory. That is, I am talking about studies of such phenomena as program effectiveness, in which an attempt is made to explain why the students in some programs achieve more than the students in others. This is in contrast to descriptive studies like National Assessment, which attempt to describe such things as the levels of achievement obtained in various sectors of American schooling. As Jöreskog (1976) put it, "a descriptive study involves the collection of data, tabulation and graphical presentation of statistical facts, whereas, an explanatory study also involves the specification of causal hypotheses,

estimation of causal models and testing of the validity of these models and hypotheses" (p. 1).

In 1956, Wold observed that the methodology for explanatory observational studies was a vastly underdeveloped field of inquiry, particularly among statisticians. He observed that most of the work that was being done in the area was being done by people in fields such as economics, genetics, and sociology. Although there have been some important improvements in the methodology for such studies since Wold's review, the development and application of this methodology has received far too little attention in educational research.

There are many reasons why this is so. One important factor has been the dominance of experimental paradigms in educational research. Many of us have been taught to believe that the only possible way of supporting causal inferences is through the use of true experimental designs. If randomization is not possible, then we are to conduct a quasi-experiment, realizing that we will suffer some loss of internal validity. However, it is important to remember that most of what is known about people and the universe has not been based on experimentation, but on observation. If we are to further our understanding of educational processes, and do so in a way that will allow us to improve those processes, it is essential that we give more attention to developing methods for conducting explanatory observational studies.

At times you may think that I am coming out against the use of experiments. I want to assure you tha this is not the case. Experimentation is

clearly a powerful scientific procedure. But we must recognize that experimental paradigms are not the only ones available, and that explanatory observational studies have the potential to make major contributions to our understanding of educational processes.

It is also important to point out that I am talking about quantitative observational studies. Admittedly, there has been only limited success in quantifying even those things we all "know" to be true (for instance, that school differences make a difference). However, I strongly believe that quantitative approaches are a necessary ingredient in a total effort toward achieving an understanding of educational processes (Cooley, Leinhardt, & McGrail, 1977).

A major problem with observational studies has been that they are generally defined negatively. For example, reference is made to the fact that randomization is absent, or treatment is not controlled, or one is applying non-experimental designs, etc. It usually sounds apologetic. What I want to do today is focus on the major requirements for sound observational studies, requirements that, if met, can yield consistent, convincing, useful explanations of educational phenomena. I shall discuss what I consider to be the most critical features of such studies under the following three topics:

(1) the sampling framework, which affects the generalizability of the observed relationships;

(2) the theoretical model, which describes the hypothesized causal structure of the variables under consideration; and

(3) the statistical procedure, which is used to analyze the network of observed relationships for the purpose of establishing the plausibility of the theoretical model and estimating its parameters.

The Sampling Framework

First, let us consider the question of population and its relevance to observational studies. There seem to be "two branches" in the literature on observational studies. One branch, exemplified by the writings of Cochran (1965) and others (e.g., Lord, 1960; Kenny, 1975; Overall & Woodward, 1975), deals with multi-population observational studies, in which different samples are administered different treatments, but random assignment of subjects to treatments is absent. I consider this branch as part of the literature on quasi-experimental design.

Others (e.g., Wold, 1956; Wiley & Hornik, 1973) consider observational studies that involve a single population. In such studies, one defines a population, draws a sample from it, and studies the relationships among variables measured on that sample. For random samples from a population, the errors in estimating relationships for the population are a function of the sample size, the reliability with which the variables can be measured, and the degree to which measures can be obtained for the entire sample. Important contributions of methodologists have been to show how to take the sampling error, the measurement error, and the missing data bias into account in estimating such relationships.

The estimated relationships may certainly include variables that reflect differences in educational treatments, but in single population studies the treatments are observed as they are currently occurring in the population sampled. The question of whether one is dealing with a single population or with two or more populations is really a question of how the sample was drawn, and whether there was any planned intervention following the drawing of the sample. As Thorndike (1942) pointed out over three decades

ago, comparing two samples that have been given two treatments without random assignment to each treatment is best thought of as a comparison of two populations. Under these conditions, there are many possible differences between the two populations other than the imposed treatment differences that may be producing differences in the dependent variable.

A prime example of a multi-population study is the evaluation of the National Follow-Through Program (FT), a quasi-experiment that has attempted to examine the effectiveness of a variety of approaches to primary grade instruction for compensatory education. This study of planned variation was a quasi-experiment because it intervened in the sample selected by introducing one of 20 possible innovative programs into each school selected for participation. It was a quasi-experiment, because the new program that was used in each school was self-selected by that school, not randomly assigned to it. Intervention following sample selection changes the situation from a single population observational study (with the observed relationships generalizable to the population sampled) to a multi-population quasi-experiment. FT involves contrasts among 20 different treatment groups, each of which has its own controls, and each of which is best thought of as coming from a separate population. (See, for example, DiCostanzo & Eichelberger, 1977, for further discussion of this problem.)

In contrast to FT is the Sustaining Effects Study (SES) of compensatory education (Hoepfner, Wellisch, & Zagorski, 1977), which is currently being conducted by the System Development Corporation. This is an excellent example of a single population observational study. They have drawn a representative sample of 250 of the nation's elementary schools in a manner

that allows them to estimate population relationships between such variables as school practices and student achievement.

It may be useful to compare some important aspects of FT and SES for the purposes of illustration. Both were intended to provide information to guide federal policy in the area of compensatory education. Both involved a large number of school sites. But the FT design did not draw a representative sample of schools, nor did it randomly assign schools to treatments. Thus, it suffers from loss of both external validity (i.e., being able to generalize any observed relationships to some population of interest) and internal validity (i.e., being able to attribute student outcome differences to differences in the imposed treatments).

As SES attempts to identify effective practices for improving the development of basic skills, it will not be handicapped because it did not insure a wide variation of school practices, as the FT interventions did. That variety occurs "naturally," as can be seen in the comparison of practices found in the FT "experimental" and "control" schools. Across all participating schools, the classroom practices that were found to be related to student learning were similarly distributed in the experimental and control schools (see Cooley & Lohnes, 1976, pp. 175-180). In addition, enthusiasts of the experimental paradigm in educational research should be considerably sobered by the wide variations in program-relevant practices among classrooms within any one of the 20 FT treatments (Stallings, 1975; Abt, 1977; House et al., 1977).

The Follow Through program has been an extremely important and useful enterprise. Much has been learned from it. But the contrast between

SES and FT is most valuable because it illustrates a tendency in education to move prematurely into intervention-type, quasi-experiments before the relationships are established among the major variables currently operating in a particular causal network. Confident descriptions of such networks prior to the implementation of quasi-experiments are necessary if we hope to be able to generalize from such manipulations. Let us now turn to the need for theoretical models and return to this important point later.

The Need for Theoretical Models

So far, I have argued that if one has a well-defined population, draws a representative sample from that population, measures variables on that sample, and does not impose treatments on any portion of that sample, it is a relatively straightforward task to estimate the relationships among those variables in that population. Statistical science has developed and is continuing to improve the procedures for taking the sampling error, the measurement error, and the missing data into account in such estimations. But if we also wish to provide causal explanations in observational studies, it is also necessary to consider specification error. Here the researcher must turn from statistical considerations to subject matter considerations, where the subject matter is what is already known about the phenomenon under investigation.

Specification error occurs when a study is designed with an inadequate theoretical model. In this context, a theoretical model is a proposed causal structure, and it generally includes both established and hypothesized causal relationships among the variables under investigation. For example, specification error occurs when a variable that directly affects a dependent

variable in a causal network is not included in the set of independent variables, but is correlated with one or more of those independent variables.

It may be useful to illustrate this specification problem by referring to a favorite example of Palmer Johnson's, whom we honor annually at AERA with the P. O. Johnson Award. Professor Johnson was not exactly a humorous lecturer, but I recall his chortling a bit when he told about the almost perfect correlation that has existed over time between the number of Protestant ministers in New England, and the barrels of rum imported into Boston Harbor. He always felt that the strong relationship between those two variables should be a very sobering illustration for any graduate student who might be tempted to make causal inferences from mere correlational coefficients. My point is that anyone who would base a causal attribution on that two variable relationship is suffering from specification error. The simplest causal model for explaining that spurious correlation between rum and ministers incorporates an important third variable, the increasing size of the population of New England over that time period.

Current examples of specification error can be found in most evaluations of educational programs. For example, to evaluate a new reading program, a large school district utilized a non-randomized comparison group design to contrast their new program with the Houghton-Mifflin program, which was in more general use in the district (Webster, 1975). As is the fashion today, they used the familiar tic-tac-toe tables of O's and X's to describe their design:

$$O_1 \; X_1 \; O_2$$
$$O_1 \; X_2 \; O_2$$

where X_1 was the old reading program, X_2 the new, and O_1 and O_2 represented reading pretest and posttest, respectively.

To examine whether or not this study suffered from specification error, one must ask whether there were other factors than the pretest and the gross differences between the two reading treatments that might have affected posttest differences. One such factor might have been the degree to which the students receiving the two treatments had the same opportunity to learn the knowledge and skills that were sampled in the posttest. Opportunity, in this context, is both a function of how time was allocated, and of the overlap between the curricula actually covered during the school year and the knowledges and skills sampled in the posttest. These two opportunity variables can be estimated, and they <u>are</u> very important determinants of school outcomes. When they are not incorporated into the design, there is the danger of attributing instructional effectiveness to specific programs or ways of teaching, when it is really a matter of how time was allocated, or whether the curriculum content was a good fit to the particular achievement test that was selected. (Notice that specification error is as relevant a concern in quasi-experiments as it is in observational studies.)

There are no statistical methods for guarding against specification error in explanatory observational studies. As Wold (1956) pointed out so well, this question of whether we have included the major explanatory variables in an analysis is not a statistical problem, but a question of subject-matter. Fortunately, procedures have been developed to assist the researcher in organizing the relevant subject matter into the kind of explicit theoretical model required in explanatory observational studies. Excellent recent

introductions to such procedures can be found in Duncan's (1975) Introduction to Structural Equation Models, and Heise's (1975) book on Causal Analysis.

To illustrate the role of subject matter and structural equations in the development of a theoretical model, it is helpful to consider briefly a particular subject matter. I have selected the development of the ability to read. To keep it simple, here are just three fairly safe generalizations from what is currently known about the development of this ability:

(a) reading performance varies considerably among children of any given grade or chronological age;

(b) a child's reading performance at one point in time is highly dependent upon reading performance at some prior point in time; and

(c) a child who spends more time learning to read will become a better reader than if that child spends less time learning to read.

Using these generalizations, a theoretical model that attempts to explain the development of the ability to read might begin with a simple, three-variable network as shown in Figure 1. This schematic may be expressed as a structural equation in Duncan's (1975) notation:

$$z = p_{zy}\, y + p_{zx}\, x + v$$

In the language of structural models, x and y are causally prior to z. The disturbance term v represents the other factors that affect reading performance (z), which are not specified in this model. The variables x and y are exogenous in this model (i.e., their values are not explained by the model), but x and y may be intercorrelated (thus the curved, double-headed arrow), but are assumed to be uncorrelated with v. The structural coefficient (p_{zx})

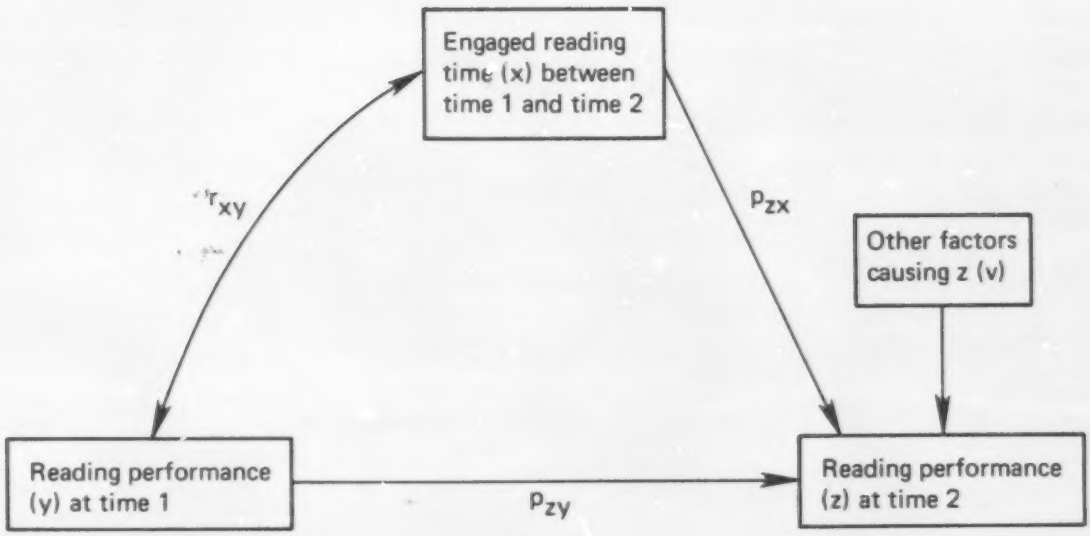
Engaged reading time (x) between time 1 and time 2
r_{xy}
p_{zx}
Other factors causing z (v)
Reading performance (y) at time 1
p_{zy}
Reading performance (z) at time 2

Figure 1. A really simple causal diagram for explaining performance on a reading test.

indicates the amount of change in z that can be expected from a change in x for a given y. Expressing these variables in standard score form,* the algebra of structural equations reveal that

$$p_{zy} = \frac{\rho_{zy} - \rho_{zx}\rho_{xy}}{1 - \rho^2_{xy}}$$

and

$$p_{zx} = \frac{\rho_{zx} - \rho_{zy}\rho_{xy}}{1 - \rho^2_{xy}}$$

where the ρ's are the corresponding population correlations. Thus, for a given population the structural coefficients for x and y are simply a function of the correlations among these three variables.

For example, if in the population of third graders, $\rho_{zy} = .60$, $\rho_{zx} = .30$ and $\rho_{xy} = .20$, then the structural equation becomes:

$$z = .56\ y + .19\ x + v$$

This suggests that for a given value of y, an increase of one standard deviation in reading time (x) will result in a .19 standard deviation increase in reading performance, assuming, among other things, that the model is correctly specified and that these correlations for the third grade population are known. Let me illustrate the importance of the latter assumption first. Those of you familiar with third grade reading research may agree that the correlation between fall pretest and spring posttest is about .60, and the other two correlations are probably between .2 and .3. But is

---

*Not necessary, but desirable for ease of illustration.

time more highly correlated with y or z? It makes a big difference. For example, if we reverse the .20 and .30, the structural coefficient for time goes from an interesting .19 to a dismal .02. This difference shows the importance of having good estimates of the population relationships.

Returning to the first assumption, which is that Figure 1 represents an adequate model of reading achievement, we encounter even greater problems. Although you may be convinced that a child has a better chance of becoming a better reader in reading class than while playing baseball, you may also believe that for a child with a given pretest (y), the effect of engaged time (x) surely depends on whether the time is spent trying to read the Encyclopedia Britannica or in one-to-one interaction with a skilled teacher using a highly effective blending algorithm to teach word decoding.

The point is that the structural model must be consistent with what is currently known about the phenomenon being modeled. For example, a given model is inadequately specified if an examination of the relevant subject matter reveals an additional variable that is known to be related functionally to reading achievement, and is correlated with one of the other variables already in the model. This is because the algebra of structural equations used in this model assumes that the disturbance variable v is uncorrelated with x and y.

This latter assumption is an important point for it lends optimism to the search for finite models and points to the need to develop networks of models. For example, if research on primary grades reading suggests the importance of a motivating classroom environment in learning to read, such variables need not necessarily be part of this particular model, if it is reasonable to assume (or it can be shown) that motivating variables are relevant

because they operate through other variables in the model. To illustrate further, motivating variables may help to explain variance in actual time spent in learning to read (x), or even the correlation between x and y, but they need not be included in this particular model unless it is reasonable to assume that motivators have a direct effect on z (or an effect through some other variable not in the model but correlated with it). Of course, some degree of specification error will always be with us, but we are more likely to have a cummulatively improved understanding of educational processes if we are guided by more and more adequately specified models, and we will continue to perform non-cummulative studies if we are not guided by causal models.

Another way of looking at this specification problem is from the perspective of "strong inference." Platt (1964) emphasized the importance of being able to deal with alternative hypotheses (i.e., alternative explanations) for the phenomena we are trying to explain. What this means for observational studies is that the variables included represent possible alternatives for explaining the variance in the dependent variable. If, for example, a model for explaining reading performance does not include variables reflecting differences in home environment, the model will not be convincing unless other studies have shown that home environment is not functionally related to reading performance (which is unlikely), or that home environment operates through prior values of the dependent variable (i.e., through y in Figure 1). The latter assumption may be tenable if home environment does not change significantly between measures of y and z. (Another hypothetical possibility is that home environment does have a direct effect but is considered a component of the disturbance term v. That possibility is untenable because of the known relationship between y and home environment.)

Platt was also kind enough to suggest where strong models come from. They come from scientists who are thoroughly familiar with their subject matter. Strong models come from scientists who are also willing and able to spend a portion of each day in quiet, analytical thought, thinking through and anticipating possible challenges to their models as currently defined, and designing new studies or reanalysis of available data to meet those challenges.

To further illustrate this point, I vividly remember the time I heard that Paul Lazarsfeld had edited a book that included a section on multivariate analysis, and how amused I was to discover that it dealt with the analysis of multi-dimensional contingency tables (Lazarsfeld, Pasanella, & Rosenberg, 1972). Here I was doing elaborate, cross-lagged, multiple-partial canonical correlations involving dozens of variables, and that eminent sociologist was still messing around with chi square tables! What I did not appreciate was that his little analyses were generally more informative than my elaborate ones, because he had the "right" variables. He knew his subject matter. He was aware of the major alternative explanations that had to be guarded against and took that into account when he decided upon the four or five variables that were crucial to include. His work represented the state-of-the-art in model building, while my work represented the state-of-the-art in number crunching.

Paul Lazarsfeld was also very important in helping me think about the qualitative-quantitative distinction. His first big job was to get me out of the qualitative vs. quantitative trap. His work is living proof that it is not an either/or proposition, it is a matter of degree of emphasis. Every research study is inevitably both qualitative and quantitative to some degree. But even more important, he recognized the importance of the

interaction between qualitative and quantitative "types" in working toward an improved understanding of educational processes. We have to talk to each other and read each others work, not shout at each other. Certainly those of us who emphasize quantitative approaches can benefit greatly from those who provide rich descriptions of what happens in classrooms and schools as we work toward more convincing causal models of the processes of schooling.

Statistical Procedures

This brings us to the statistical procedures that can implement the theoretical structural model. The purpose of the statistical procedures is to assist in establishing the plausibility of the theoretical model and to estimate the degree to which the various explanatory variables seem to be influencing the dependent variables. Various forms of linear regression are generally used for this purpose. It is certainly well-known that regression analysis and analysis of covariance are formally the same statistical procedure. They are both just variations of the general linear model. What is important is not the shift from covariance analysis to regression analysis, but rather a shift to causal modeling.

But it is also important to note that the approach I am advocating also includes the shift to thinking about educational treatments as multidimensional domains, and not as distinct, discrete, homogeneous treatments worthy of consideration as levels in a typical analysis-of-variance contrast. The treatment dimensions reflect the various ways in which a program can be implemented by different teachers, as well as other important ways in which teachers and programs vary.

Of course, it is not enough to suggest that the statistical procedure be linear regression. There are many variations available (step-wise, commonality, ridge, least squares, maximum likelihood, standardized, raw, etc.), and a few well-known problems to consider, such as multicollinearity among the explanatory variables, the appropriate metric to be used to represent them, sensible levels of data aggregation, and how to deal with their errorful measurement.

In considering possible statistical models, it may be useful to distinguish between exploratory and more "confirmatory" stages in the investigation of a phenomenon. Exploratory approaches are particularly useful in current studies of the effects of educational programs because of the primitive state of relevant theoretical models. Needed at this stage are statistical procedures that allow us to see the relative usefulness of different predictors or sets of predictors, as well as what confounding is occurring among the independent variables, and what differences there are among different possible models for the data. At the exploratory stage, the data analysis should suggest ways in which the theoretical model might be modified, how the measures might be combined or separated, or which variables can safely be deleted from the model. As Lohnes and I argue elsewhere (Lohnes & Cooley, 1978), commonality analysis has been found to be useful for such exploratory data analysis. In a recent, excellent review of regression analysis, Hocking (1976) points out the possible usefulness of commonality analysis in the comparison of possible models, and "as a means of identifying the roles of the variables, both alone and in relation to other variables" (p. 8).

As more convincing models are specified, the algebra of structural equations is useful in the consideration of the appropriateness of different statistical procedures. For example, as you may have noticed, the structural coefficients for Figure 1 can be estimated as linear regression coefficients in good old multiple correlation. But it is also important to notice that I did not start with the notion that I should plug my data into a multiple regression program. The algebra of structural equation modeling showed that for the particular simple causal model of Figure 1, multiple correlation was a sensible statistical procedure for estimating the structural coefficients.

One promising set of statistical procedures for estimating the unknown structural coefficients in a set of linear structural equations has been built into a computer program called LISREL (Jöreskog & van Thillo, 1973). One advantage of LISREL is that its architects recognized the fact that the variables of primary interest are often not directly measurable, but must be estimated from measures of multiple indicators of the unmeasured, hypothetical variables (or latent factors). Another important advantage of LISREL is that it incorporates procedures for estimating the reliabilities of the variables, and good estimates of reliabilities are essential if we hope to correct for the attenuation inherent in research involving errorful measures.

The State of the Art

Having said all this, what is "the state of the art" in the design and analysis of research studies involving relationships among variables? My sense of it is that, thanks to geneticists, economists, sociologists, and others, a very exciting "art form" exists that educational research has

largely ignored. The few educational applications that have been tried or suggested (for example, Anderson & Evans, 1974; Rao, Morton, Elston, & Yee, 1977; Werts & Linn, 1970; Magidson, 1977; Werts & Hilton, 1977; McDonald & Elias, 1976; Wiley & Harnischfeger, 1973) illustrate its potential, as well as the primitive state of the causal models that have been developed, to date, for guiding such efforts.

We must clearly produce more convincing causal models. The challenge is considerable, but the efforts will surely be more productive than continuing to conduct meaningless quasi-experiments, or averaging all of the t-tests they may have produced. As educational research is guided by increasingly valid models of educational phenomena, and the causal networks that are currently operating are better understood, we may eventually draw close to that "complete covariate" that Cronbach and his associates (1976) have shown is necessary for interpretable quasi-experiments. It may then be sensible "to push entities into locations in the predictor space where they are not naturally found" (Lohnes & Cooley, 1978), and see what results.

References

Abt Associates, Inc. Education as experimentation: A planned variation model. Volume IVA. Cambridge, Mass.: Author, 1977.

Anderson, J. G., & Evans, F. B. Causal models in educational research: Recursive models. American Educational Research Journal, 1974, 11(1), 29-39.

Cochran, W. G. The planning of observational studies of human populations. Royal Statistical Society Journal, 1965, series A, 128, 234-266.

Cooley, W. W., Leinhardt, G., & McGrail, J. How to identify effective teaching. Anthropology and Education Quarterly, 1977, 8(2), 119-126.

Cooley, W. W., & Lohnes, P. R. Evaluation research in education. New York: Irvington Publishers, 1976.

Cronbach, L. J., Rogosa, D. R., Floden, R. E., & Price, G. G. Analysis of covariance: Angel of salvation, or temptress and deluder? Stanford, Calif.: Stanford University, 1976.

DiCostanzo, J. L., & Eichelberger, R. T. Design, analysis, and reporting considerations when ANCOVA-type techniques are used in evaluation settings. Paper presented at the annual meeting of the American Educational Research Association, New York, April 1977.

Duncan, O. D. Introduction to structural equation models. New York: Academic Press, 1975.

Heise, D. R. Causal Analysis. New York: John Wiley & Sons, 1975.

Hocking, R. R. The analysis and selection of variables in linear regression. Biometrics, 1976, 32(1), 2-49.

Hoepfner, R., Zaborski, H., & Wellisch, J. B. The sample for the sustaining effects study and projections of its characteristics to the national population. Santa Monica, Calif.: System Development Corporation, 1977.

House, E. R., Glass, G. V., McLean, L. D., & Walker, D. F. No simple answer: Critique of the "Follow Through" evaluation. Urbana, Ill.: University of Illinois, Center for Instructional Research and Curriculum Evaluation, 1977.

Jöreskog, K. G. Causal models in the social sciences: The need for methodological research. (Research report 76-8). Uppsala, Sweden: University of Uppsala, 1976.

Jöreskog, K. G., & van Thillo, M. LISREL: A general computer program for estimating a linear structural equation system involving multiple indicators of unmeasured variables. (Research report 73-5). Uppsala, Sweden: University of Uppsala, 1973.

Kenny, D. A. A quasi-experimental approach to assessing treatment effects in nonequivalent control group design. Psychological Bulletin, 1975, 82, 345-362.

Lazarsfeld, P. F., Pasanella, A. K., & Rosenberg, M. Continuities in the language of social research. New York: The Free Press, 1972.

Lohnes, P. R., & Cooley, W. W. Regarding criticisms of commonality analysis. Paper presented at the annual meeting of the American Educational Research Association, Toronto, Canada, March 1978.

Lord, F. M. Large-scale covariance analysis when the control variable is fallible. Journal of the American Statistical Association, 1960, 55, 307-321.

Magidson, J. Toward a causal model approach for adjusting for preexisting differences in the nonequivalent control group situation: A general alternative to ANCOVA. Evaluation Quarterly, 1977, 1(3), 399-420.

McDonald, F. J., & Elias, P. Beginning teacher evaluation study, phase II. Executive summary report, 1973-74. Princeton, N. J.: Educational Testing Service, 1976.

Overall, J. E., & Woodward, J. A. Nonrandom assignment and the analysis of covariance. Galveston, Texas: University of Texas, Psychometric Laboratory, 1975.

Platt, J. R. Strong inference. Science, 1964, 146, 347-353.

Rao, D. C., Morton, N. E., Elston, R. C., & Yee, S. Causal analysis of academic performance. Behavior Genetics, 1977, 7(2), 147-159.

Ross, L., & Cronbach, L. J. Review of Handbook of evaluation research by M. Guttentag & E. L. Struening (Eds.). Educational Researcher, 1976, 5(10), 9-19.

Stallings, J. Implementation and child effects of teaching practices in Follow Through classrooms. Monographs of the Society for Research in Child Development, 1975, 40 (7-8, Serial No. 163).

Thorndike, R. L. Regression fallacies in the matched groups experiment. Psychometrika, 1942, 7, 85-102.

Webster, W. J. Abstracts of research and evaluation report 1974-75. Dallas, Texas: Dallas Independent School District, 1975.

Werts, C. E., & Linn, R. L. Path analysis: Psychological examples. Psychological Bulletin, 1970, 74(3), 193-212.

Werts, C. E., & Hilton, T. L. Intellectual status and intellectual growth, again. American Educational Research Journal, 1977, 14(2), 137-146.

Wiley, D. E., & Harnischfeger, A. Post hoc, ergo propter hoc: Problems in the attribution of change. (Report No. 7). Chicago: University of Chicago, 1973.

Wiley, D. E., & Hornik, R. Measurement error and the analysis of panel data. (Report No. 5). Chicago: University of Chicago, 1973.

Wold, H. Causal inferences from observational data: A review of ends and means. Journal of the Royal Statistical Society, Series A, 1956, 119, 351-390. Reprinted in M. C. Wittrock, & D. E. Wiley (Eds.), The evaluation of instruction: Issues and problems. New York: Holt, Rinehardt and Winston, 1970.